

Review Article

***In silico* Prediction of Immunogenic T Cell Epitopes of *Leishmania donovani donovani* GP63 Protein: an Alternative Approach for Anti-parasite Vaccine Development**

Mona E. E. Elfaki ¹, Anne S. De Groot ^{2,3}, Andres H. Gutierrez ², Brema M. Younis ¹, Rayan Tassone ³, Francis Terry ³, Ahmed M. Musa ¹, Ahmed M. Elhassan ¹, Eltahir A. G. Khalil ^{1*}

¹Department of Clinical Pathology and Immunology, Institute of Endemic Diseases, University of Khartoum, Khartoum, Sudan.

²Institute for Immunology & Informatics, University of Rhode Island, Providence, Rhode Island, USA.

³EpiVax, Inc., Providence, Rhode Island, USA.

*Corresponding author: Professor Eltahir Awad Gasim Khalil, Department of Clinical Pathology, Institute of Endemic Diseases, University of Khartoum, P.O. Box 45235, Khartoum 11111, Sudan. Email: eltahir@iend.org eltahirgasim@yahoo.ca

Received: 06-09-2015

Accepted: 07-10-2015

Published: 08-04-2015

Copyright: © 2015 Eltahir

Abstract

Visceral leishmaniasis (VL) is a major parasitic childhood disease in sub-Saharan Africa. Expensive and toxic anti-leishmanial drugs are current control methods. Safe, effective and cheap vaccines are potentially powerful strategies to control VL. Traditional vaccine development techniques have failed to deliver an effective vaccine. Leishmania vaccine development may benefit from immunoinformatics tools. This paper describes an improved *in silico* prediction method for immunogenic *Leishmania donovani donovani* GP63 protein T cell epitopes as VL candidate vaccines. Using the EpiMatrix algorithm, the amino acid sequence of *Leishmania Donovani donovani* GP63 protein (GenBank accession: ACT31401) was screened for putative T cell cluster epitopes that would bind to the most common HLA class II alleles among at-risk populations. Nine epitopes were initially identified using EpiMatrix. Based on cluster score, number of EpiMatrix hits, hydrophobicity, and number of EpiBars (an EpiBar is a 9 amino acid frame predicted to bind at least 4 different HLA molecules), four peptides (P1-P4) were selected for synthesis. In a proof of concept study, blood samples from consenting healthy, leishmanin skin test (LST) reactive and non-reactive volunteers were stimulated and IFN- γ , IL-4, and IL-10 were measured. IFN- γ and IL-4 levels were similar in both groups. However, mean IL-10 levels were significantly reduced in LST reactive individuals. To evaluate whether cross-reactivity with the human genome (HG), the human gut microbiome (HM) and common human pathogens (HP) was responsible for these differences, the sequences of the evaluated peptides were screened using JanusMatrix. One of the peptides (P1), which increased IL-10 in the LST reactive volunteers, showed high cross-reactivity with HG, suggesting that P1 might induce a regulatory immune response in humans. In conclusion, immunoinformatics tools provide a promising alternative approach for anti-parasite vaccine development. Data obtained can be used in the development of epitope-based *Leishmania* vaccine.

Keywords: HLA Class II Alleles; Regulatory T Cell Epitope; Visceral Leishmaniasis; Epimatrix

Introduction

Visceral leishmaniasis (VL) is recognized as a major public health problem in a number of countries spanning four continents, with an incidence of 0.5 million new cases per year and considerable fatalities. Currently, prevention of VL focuses on case detection/treatment, vector and animal reservoir control and vaccine development. Treatment options are limited, however; no effective vaccine has been developed to date [1-6]. Safe, effective and cheap vaccines are critically important for the effective control of VL. It is likely that a vaccine can be developed for leishmaniasis because spontaneous or drug-induced recovery from cutaneous or visceral leishmaniasis is associated with solid and life-long immunity against re-infection. First-generation and available second-generation vaccines have failed to deliver a candidate that was successful in Phase III trials. Ideally, a vaccine for VL should elicit a strong Th1 response and little or no Th2 response [7-11]. Immunoinformatics tools that analyze the sequence of parasite proteins for binding to common human leukocyte antigen (HLA) class II alleles can provide an expedited approach to vaccine development [12].

Glycoprotein-63 (GP63; Leishmanolysin) has been implicated in various aspects of the host-parasite relationship. It functions as a receptor for host macrophages, converting complement component C3b to C3bi in order to facilitate parasite survival. GP63 is able to cleave multiple intracellular signaling proteins of the host, contributing to the infectivity of *Leishmania* [13,14]. In mouse models, GP63 vaccination has been shown to induce protective immune responses against leishmaniasis through induction of antigen-specific T cells [15,16]. GP63 is highly conserved among all species of *Leishmania* and is the major surface protein expressed in both promastigote and amastigote life stages [17-20].

Protective immune responses to leishmaniasis depend entirely on Th1 cells, which produce IFN- γ and other Th1 cytokines [21]. The activation of CD4+ helper T cells is essential for the development of adaptive immunity against intracellular pathogens [22-25]. A critical step in CD4+ T cell activation is the recognition of epitopes presented by HLA class II molecules [26]. HLA class II molecules are heterodimers expressed on the surface of professional antigen presenting cells that bind linear peptide fragments derived from protein antigens [27]. A hallmark of the HLA class II binding peptide groove is that there are four major pockets. These pockets accommodate side chains of residues 1, 4, 6, and 9 of a 9-mer core region of the binding peptide; this core region interaction largely determines binding affinity and specificity [28]. In addition, peptide residues immediately flanking the core region have been shown to make contact with the HLA molecule outside of the binding groove and to contribute to HLA-peptide interaction [29].

Binding motifs derived for HLA class II molecules are highly variable. An epitope's HLA binding promiscuity is an im-

portant feature for vaccine development and immunotherapy due to its ability to bind multiple HLA class II molecules (alleles) and the population coverage conferred by targeting multiple alleles at once [30, 31]. The goal of T cell epitope prediction is therefore to accurately identify peptide sequences within any protein that, in the context of a defined HLA molecule or across multiple families of HLA haplotypes, will elicit desired T cell responses when presented at the cell surface to be recognized by the immune system. These epitopes should ideally be dominant and promiscuous, so that they are recognized by most human populations [32]. Immunoinformatics tools significantly accelerate the discovery of peptide epitope sequences, which account for an overwhelmingly large proportion of T cell responses to pathogens and self-proteins [12,32]. One such tool is EpiMatrix, a T cell epitope mapping algorithm developed by De Groot and Martin (EpiVax). EpiMatrix screens protein sequences for 9–10 amino acid long peptide segments predicted to bind to one or more MHC alleles [33,34]. EpiMatrix uses the pocket profile method to build matrices for epitope prediction, which was first described by Sturniolo and Hammer [35]. EpiMatrix has been successfully applied to the analysis of previously published epitopes [36]. For example, researchers affiliated with GAIA (Global Alliance to Immunize against AIDS) have used EpiMatrix to identify highly conserved and immunogenic epitopes from available HIV sequences [37-40]. EpiMatrix was also used to identify epitopes for inclusion in novel, broadly reactive epitope-based vaccines for *F. tularensis* [41], *M. tuberculosis* [42], vaccinia virus [43, 44], *H. pylori* [45] and Influenza A (H1N1) virus [46,47]. In this communication, we describe the application of EpiMatrix and JanusMatrix, a new tool that predicts pathogen-host cross-reactivity at the T cell receptor level, to scan a parasite surface protein sequence (*Leishmania donovani donovani* GP63 protein, GenBank accession: ACT31401) for T cell epitopes that will bind the most common HLA-DRB1 alleles among at-risk populations.

We also describe the application of JanusMatrix to examine potential cross-reactivity of the peptides with human genome (HG), the human gut microbiome (HM), and common human pathogens (HP). JanusMatrix describes patterns of HG, HM, and HP cross-reactivity (XR) (XR is the number of cross-reactive hits detected) that are distinct for epitopes associated with regulatory T cell and effector T cell responses [48]. Greater XR with HG compared to HM seems to distinguish defined regulatory and effector T cell epitopes [48,49]. JanusMatrix was used to refine our in vitro observations since it provides some additional insight into the phenotype of T cell responses that may occur. The application of these bioinformatics tools represents an alternative approach to identify possible candidates for anti-leishmania vaccine development.

Results

HLA-DRB1 allele frequencies

A total of eleven HLA-DRB1 allelic groups assessed among

the studied population from the Institute of Endemic Diseases Database, Khartoum, Sudan. The most common HLA-DRB1 alleles within the study population were DRB1*1101 and DRB1*0804, with frequencies of 32.4% and 24.1%, respectively [50].

Secretion analysis for the *Leishmania donovani donovani* GP63 protein

GP63 protein was predicted to have a signal peptide (1-39), a glycosylphosphatidylinositol (GPI) anchor and no trans-membrane regions or lipoprotein attachment sites, confirming that it is a surface protein [51].

EpiMatrix and ClustiMer Analysis results

Using EpiMatrix predictive matrices for HLA DRB1*0804 and DRB1*1101, the immunogenicity score of whole GP63 was calculated to be -55.83 (Fig. 1a); no clusters were identified, as only two alleles were scored. However, some 9-mers were found with scores above 1.64 (Table 1). Accounting for the set of eight common HLA-DR super type alleles, the GP63 protein immunogenicity score was -41.35. Although this score was very low, indicating that the whole protein is unlikely to be very immunogenic (Figure 1), ClustiMer analysis detected several putative T cell epitope clusters that may be useful to drive immune response. The cluster with the highest score (25.29) was within the signal peptide of the GP63 protein (Table 1).

Using EpiMatrix predictive matrices for the set of eight common HLA-DR super type alleles, nine clusters were identified in the GP63 protein (Table 1). The cluster addresses (Table 1) indicate the location of the peptides within the GP63 sequence. The core peptide (underlined middle amino acids in bold) defines the actual epitope cluster containing the highest binding probability. The flanking residues (N-terminal and C-terminal, in italics) are included to stabilize the synthesized core sequence in the open-ended binding groove and to enhance TCR recognition [52]. The EpiMatrix cluster score was derived from the deviation from the number of predicted ligands expected by random chance in a peptide of the same length (Table 1).

Four clusters were synthesized for this study: cluster 7 (*STHRHRSVAARLVRLAAAGAAVIA*); cluster 151 (*DILVKHLIPQALQLH*); cluster 205 (*TDFVYMYVASVPSEGDVL*); and cluster 496 (*SHGIUKSYAGLCANVRCDT*) (Table 1). Selection was based on the EpiMatrix cluster score, number of EpiMatrix hits, hydrophobicity, number of EpiBars and homology to human proteins. Cluster 7 (cluster address 7-30) was selected for synthesis because it had the highest EpiMatrix cluster score of 25.29 as well as three EpiBars and 17 EpiMatrix hits. In addition, it was identified as part of the signal peptide with no significant homology to human proteins by BLAST. Cluster 7 is expected to be highly immunogenic as compared to well-known immunogens.

Figure 1a

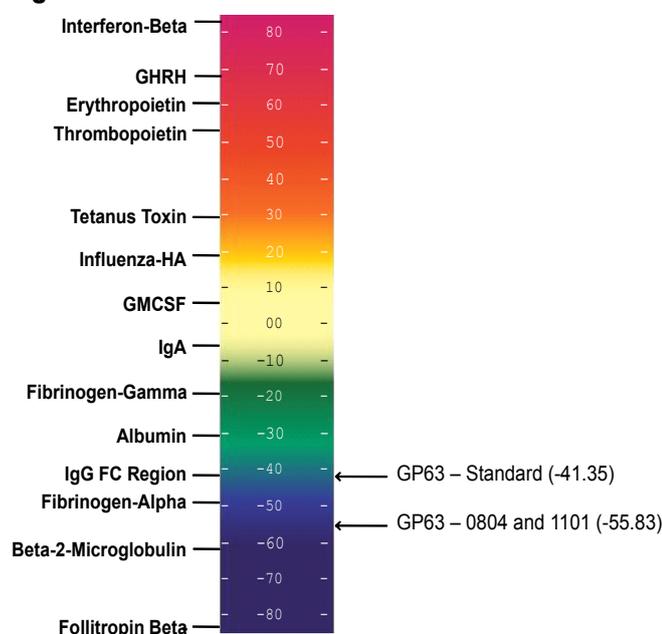


Figure 1b

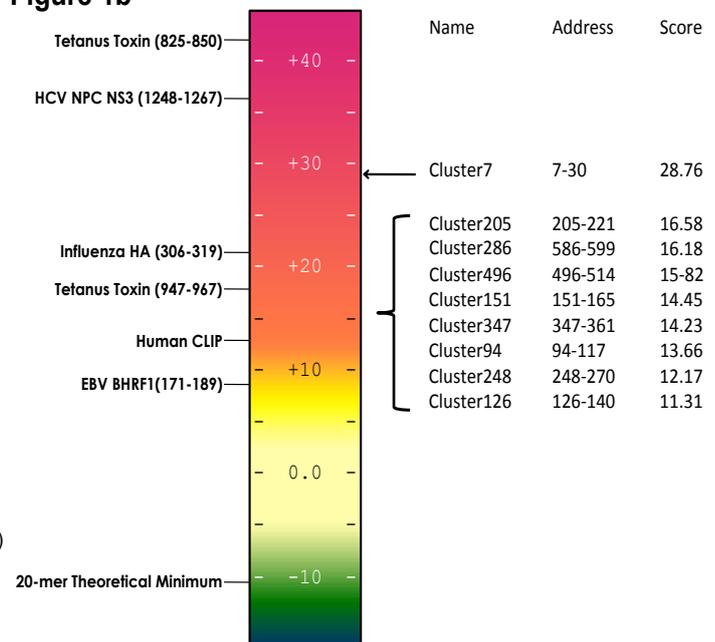


Figure 1. Immunogenicity Scale Report. EpiMatrix was used to score GP63 for putative binding to eight common HLA alleles (GP63 – Standard) and to DRB1*0804 and DRB1*1101, the two most commonly found class II HLA alleles among VL at-risk populations (GP63 – 0804 and 1101); scores are displayed in Fig. 1a as compared to standard immunogens. GP63 was then scanned for immunogenic clusters using ClustiMer; clusters identified in this analysis are displayed in Fig. 1b as compared to standard immunogenic clusters.

Table 1. *Leishmania donovani donovani* GP63 protein (GeneBank accession: ACT31401) clusters identified by EpiMatrix and ClustiMer analysis.

ClusterAddress (w/FLANKS)	Cluster Sequence	Hydrophobicity	EpiMatrix HITS (w/o FLANKS)	EpiMatrix CLUSTER SCORE (w/o FLANKS)	EpiBars
7 – 30*	STH <u>RHR</u> SVAARLVRLAAAGAAVIA	0.5	17	25.29	Yes (3)
94 – 117	APT <u>VV</u> RAANW GALRIA VSTEDLTD	0.2	11	13.66	No
126 – 140	GQR <u>IK</u> HALGRFAICT	0.04	6	11.31	Yes (1)
151 – 165*	DIL <u>VK</u> HLIPQALQLH	0.52	7	14.45	Yes (1)
205 – 221*	TDF <u>V</u> MYVASVPSEGDVL	0.65	10	16.58	Yes (2)
248 – 270	AS <u>RY</u> DQLVTRVVTHEMAHALGFS	-0.01	10	12.17	Yes (2)
347 – 361	AAG <u>Y</u> SALTMALFQD	0.58	7	14.23	Yes (1)
496 – 514*	SHG <u>I</u> KS YAGLCANVRCDT	0.14	10	15.82	Yes (2)
586 – 599	ATA <u>LL</u> VAAALLAVAL	2.7	10	16.18	Yes (1)

*Cluster addresses for the predicted peptides P1, P2, P3 and P4.

Clusters 205 (cluster address 205-221) and 496 (cluster address 496-514) were selected because both have EpiMatrix cluster scores >15 (16.58 and 15.82, respectively), and each contains two EpiBars and ten EpiMatrix hits. These clusters showed no significant homology to human proteins based on BLAST analyses and compared well to standard immunogens (Figure 1b). Clusters 151 (cluster address 151-165) and 347 (cluster address 347-361) had almost the same EpiMatrix cluster scores (14.45 and 14.23, respectively), similar numbers of EpiBars, and similar number of EpiMatrix hits. However, Cluster 151 was chosen for synthesis because it included a top 1% EpiMatrix hit (Z-score > 2.32). This hit is specific for DRB1*0804, one of the two alleles associated with at-risk populations.

Clusters 94 (cluster address 94-117), 126 (cluster address 126-140) and 248 (cluster address 248-270) were excluded from selection because they have comparatively low EpiMatrix cluster scores (13.66, 11.31 and 12.17, respectively). Cluster 586 (cluster address 586-599) had an EpiMatrix cluster score of 16.18 placing it relatively high on the immunogenicity scale (Figure 1), but it was excluded due to elevated hydrophobicity (Table 1) and identity with sequences in the human genome (70%). None of the other predicted clusters had any significant linear homology to the human genome based on BLAST analysis. Many of the clusters were highly conserved among the other *Leishmania* species, specifically those four clusters/epitopes selected for synthesis. The predicted and synthesized peptides were assigned the following numbered letters: P1, P2, P3 and P4 for the sake of simplicity [50].

P1: STHRHRSVAARLVRLAAAGAAVIA (Cluster 7)

P2: DILVKHLIPQALQLH (Cluster 151)

P3: TDFVMYVASVPSEGDVL (Cluster 205)

P4: SHGIKS YAGLCANVRCDT (Cluster 496)

***In vitro* whole blood stimulation results**

The *in vitro* immunogenicity studies for the predicted peptides showed that IFN- γ was not significantly increased in any of the volunteer groups. The peptide pool (P1+P2+P3+P4) produced a moderate IFN- γ increase that was not statistically significant in LST non-reactive volunteers. IL4- was not significantly increased in any group (LST-reactive= skin induration \geq 5mm; LST-non reactive= skin induration = 00 mm). IL-10 was significantly increased following stimulation with the P1 peptide in all volunteers, especially in the LST-reactive group. Significant IL-10 production was observed following stimulation with the P3 peptide in LST non-reactive volunteers. The peptide pool significantly decreased IL-10 production in all volunteers, especially in the LST-reactive group.

Janus Matrix analysis results:

Despite its lack of homology based on BLAST analyses, the P1 peptide was found to be highly cross-reactive with the HG (XR=33, Ratio=2.92) using JanusMatrix, suggesting that P1 may be a regulatory T cell epitope (Table 2). Furthermore, one 9-mer epitope within P1 had 14 of the 33 cross-

reactive hits (Fig. 2) with HG (Ratio = 1.24). This contrasts with P4, which has a low XR ratio (0.35) and induced only low levels of IL-10; P4 appears to be an effector T cell epitope. P2 and P3 showed a minimal degree of cross-reactivity with the HG, HM and HP (Table 2).

trix); blue triangles are 9-mers that are 100% identical to the TCR face of the source epitope and that are predicted to bind to the identical HLA; and light purple circles are proteins containing the cross-reactive epitope.

Table 2. Cross-reactivity of the predicted peptides with the Human Genome (HG), Human Microbiome (HM) and Human Pathogens (HP).

Source	Type of Response	Cluster score	HG			HM		HP		LST-reactive (n=10)	LST- non reactive (n=12)	Total reactive
			XR ^a	Ratio ^b	Janus Human Score ^c	XR	Ratio	XR	Ratio			
P1	Increased IL-10 production	25.29	33	2.92	3.11	405	1.85	586	4.00	7/10 [70%]	2/12 [16.7%]	9/22 [40.9%]
P2	None	14.45	2	0.18	0.20	49	0.22	40	0.27	0/10 [0%]	0/12 [0%]	0/22 [0%]
P3	Increased IL-10 production	16.58	2	0.18	0.27	92	0.42	48	0.33	4/10 [40%]	10/12 [83.3%]	14/22 [63.6%]
P4	Decreased IL-10 production	15.82	4	0.35	0.50	59	0.27	30	0.20	9/10 [90%]	5/12 [41.7%]	14/22 [63.6%]

XR^a is the number of cross-reactive hits detected.

Ratio^b is the ratio of cross-reactive hits per 1×10^6 amino acids in the comparison database.

Janus Human Score^c represents the average depth of coverage in the search database for each EpiMatrix hit in the input se

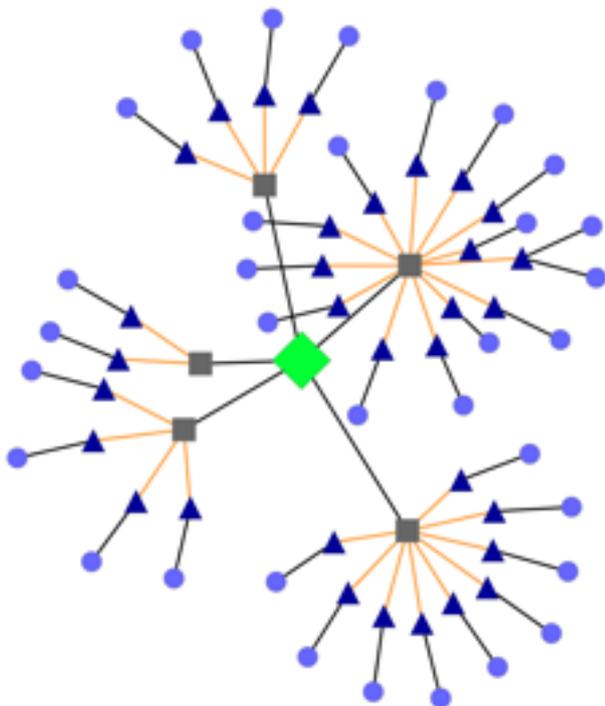


Figure 2. TCR-Epitope network for P1 peptide. Epitopes with TCR-facing residues identical to P1 were identified in protein sequences from the human genome database. Green diamonds represent P1 (source peptide); gray squares are predicted 9-mer epitopes derived from the source peptide (predicted using EpiMa-

Discussion

First- and second-generation anti-parasite vaccines against leishmaniasis have not yet shown long-term protective potential. This could be attributed to a number of obstacles, including low immunogenicity, lack of dose standardization and variability in evaluating the immune responses in challenged individuals [53, 54]. Alternative approaches for anti-parasite vaccine development are urgently needed. Recently, T cell epitope prediction techniques using advanced immunoinformatics tools have provided real hope as a new strategy for subunit vaccine development. Common HLA class II alleles of at risk-populations can be a good starting point for T cell epitope prediction. A major step in identifying potential T cell epitopes involves identifying the peptides that bind to a target major HLA molecule. Because of the high cost of experimental identification of such peptides, computational methods for predicting HLA binding peptides are highly useful, as they save both time and money.

Compared to other vaccine types, synthetic peptide vaccines do not carry whole gene sequences, can be standardized, multiple epitopes accounting for variants of a pathogen can be included and vaccines can be developed for pathogens that are difficult to culture in the laboratory. In addition, an appropriate peptide-based vaccine would also decrease the chance of stimulating a response against self-antigens. Poor

immunogenicity, necessitating the use of potent adjuvants, is a potential problem that arises with short peptide vaccines [55].

High-frequency HLA class II alleles of at risk populations (DRB1*0804 and DRB1*1101) were used to predict potential immunogenic peptides of *Leishmania donovani donovani* GP63 protein.

To confirm the immunogenicity of the predicted peptides, a small study was conducted [50]. The study concluded that the synthetic peptides (P1-P4) produced no clear Th1-response individually. The peptide pool marginally increased IFN- γ secretion in whole blood assays, suggesting that they could be the basis of a subunit vaccine in combination with a potent adjuvant. There was no significant IL-4 production upon peptide stimulation. Two peptides (P1 and P3) were observed to induce strong IL-10 up-regulatory effects in vitro.

The antigen-presenting pathway is constantly processing proteins from infectious organisms and exposes peptides derived from these proteins to the immune system at the cell surface bound to HLA molecules. One way for infectious organisms to camouflage themselves and avoid immune recognition is to either reduce HLA binding (deimmunization) and/or to modify the TCR profile of their constituent peptides to resemble host sequences [P1 peptide] (tolerization). P4 and the peptide pool have marked IL-10 reduction effects, making them probable immune therapies for infections with increased IL-10 levels and good candidates for potential inclusion in a *L. donovani* vaccine based on their high putative immunogenicity (Figure 1b) [48, 56-59].

These first studies show significant differences in immune responses to specific peptides in patients who had been exposed to *L. donovani* infections. The phenotype of the response correlates with JanusMatrix results, at least for IL-10 production in LST-reactive individuals.

Conclusion

Immunoinformatics analysis of the potential MHC-binding affinity of parasite surface protein sequences to common HLA class II DRB1 molecules provides a useful alternative method for anti-parasite vaccine development and predicts the phenotype of their immune responses.

Methods

Ethics Statement

The study protocol was scientifically and ethically reviewed and approved by the Scientific and Ethics Committees of the Institute of Endemic Diseases, University of Khartoum. Written informed consent was obtained from all volunteers for the pilot study.

Common HLA-DRB1 alleles among population at risk

The most common HLA-DRB1 alleles of patients with para-

sitologically confirmed VL/post kala-azar dermal leishmaniasis (PKDL) patients, apparently healthy volunteers with and without reactive leishmanin skin test (LST) and without history of VL/PKDL were obtained from the Institute of Endemic Diseases database. Volunteers were excluded if they failed to give written informed consent or if they have blood relation to any the enrolled patients or volunteers [50].

Leishmania donovani donovani GP63 protein sequence analysis

Standard analysis of the sequence of the *Leishmania donovani donovani* GP63 protein was performed as previously described [51]. *Leishmania donovani donovani* GP63 protein (GenBank accession: ACT31401) was analyzed using Phobius and LipoP servers. Phobius was used to identify signal peptides and transmembrane segments [60], while LipoP was used to identify lipoprotein attachment sites in the protein [61]. Presence of glycosylphosphatidylinositol (GPI) anchors was evaluated as well using GPI-SOM algorithm available at ExPASy [62]. Each 9-mer frame (9-amino acid sequence) was assessed using EpiMatrix [63] to determine binding affinities to the most commonly found class II HLA alleles among VL at-risk populations (DRB1*0804 and DRB1*1101) and to eight common HLA alleles that cover >90% of human populations worldwide (DRB1*0101, DRB1*0301, DRB1*0401, DRB1*0701, DRB1*0801, DRB1*1101, DRB1*1301, and DRB1*1501). Each 9-mer frame was assigned a Z-score ranging from approximately -3 to +3. Z-scores ≥ 1.64 , generally comprising the top 5% of any given peptide set, were defined as "Hits" and considered potentially immunogenic. Z-scores ≥ 2.32 comprise the top 1% and were considered extremely likely to bind the HLA molecule for which the score was assigned. A 9-mer frame predicted to bind to at least 4 different HLA alleles is termed an EpiBar. EpiBars may be the signature feature of highly immunogenic, promiscuous class II epitopes. For the analysis of the eight common HLA alleles, regions of high epitope density, or clusters, were identified and scored by the ClustiMer algorithm. An ancillary algorithm, ClustiMer identifies "clustered" or promiscuous epitopes [40, 41]. ClustiMer was also utilized to perform a grand average of hydropathicity (GRAVY) analysis, which identified epitopes with elevated hydrophobicity, triage of which can minimize technical difficulties with peptide synthesis and low water solubility stemming. BlastMer identifies homologies between the putative epitopes identified by EpiMatrix and any protein sequence on file at GenBank. EpiMatrix cluster score and EpiMatrix whole protein score were derived from the number of hits normalized for the length of the cluster or whole protein. The EpiMatrix cluster score is a representation of predicted promiscuity across multiple HLA alleles; clusters scoring above 10 are considered potentially immunogenic and therefore good candidates for T cell epitope-based vaccines. The predicted epitopes were screened for immunogenicity on a scale that compares the protein to vaccine antigens and other biologics already in clinical use (Figure 1).

The pilot study: In vitro whole-blood assays on healthy and LST reactive volunteers

Four of the predicted epitopes were selected for synthesis by 21st Century Biochemicals (Marlborough, MA, USA) and tested *in vitro* in whole blood samples from 22 volunteers as a proof of principle. Twelve volunteers were LST non-reactive (induration = 0 mm), while the rest (n = 10) were LST reactive with a mean LST induration of 12.6 ± 6 mm. Cytokine ELISA was used to measure IFN- γ , IL-4 and IL-10 as per manufacturer's instruction (Komabiotech Inc., Seoul, South Korea) [50].

Janus Matrix Analysis

Peptide sequences of P1-P4 were analyzed using JanusMatrix to identify potential cross-reactive TCR-facing epitopes with HG, HM, and HP databases. Ratios of cross-reactive hits (number of cross-reactive hits divided by the total number of amino acids in the comparison database multiplied by 1 x 10⁶) per cluster were calculated to determine potential regulatory T cell epitopes within the clusters. In addition, JanusMatrix results were aggregated into a Janus Human Score for each cluster, which represents the average depth of coverage in the search database for each EpiMatrix hit in the input sequence. For example, an input peptide with eight EpiMatrix hits, all of which have one match in the search database, has a Janus Homology Score of 1. An input peptide with four EpiMatrix Hits, all of which have two matches in the search database, has a Janus Homology Score of 2. Finally, epitope network diagrams were created from JanusMatrix output using Cytoscape (Fig. 2) [64] to visualize the cross-reactivity between the peptides and the target genome databases [48].

References

- Alvar J, Vélez ID, Bern C, Herrero M, Desjeux P et al. Leishmaniasis worldwide and global estimates of its incidence. *PLoS ONE*. 2012, 7(5): e35671.
- Zijlstra EE and El-Hassan AM. Leishmaniasis in Sudan. Post kala-azar dermal leishmaniasis. *Transactions of the Royal Society of Tropical Medicine*. 2001, 95(1): 59-76.
- Desjeux P. Leishmaniasis: current situation and new perspectives. *Comparative Immunology, Microbiology and Infectious Diseases*. 2004, 27(5): 305-318.
- Mathers CD, Ezza Mt and Lopez AD. Measuring the burden of neglected tropical diseases: the global burden of Disease framework. *PLoS Neglected Tropical Diseases*. 2007, 1(2): 114.
- World Health Organization. The Global Burden of Disease: 2004 update. WHO, Geneva, Switzerland. 2008.
- Hailu A, Musa A, Wasunna M, Balasegaram M, Yifru S et al. Geographical variation in the response of visceral leishmaniasis to paromomycin in East Africa: a multicentre, open-label, randomized trial. *PLoS Neglected Tropical Diseases*. 2010, 4(10): 709.
- Khalil EAG, Zijlstra EE, Kager PA and El-Hassan AM. Epidemiology and clinical manifestations of Leishmania donovani infection in two villages in an endemic area in eastern Sudan. *Tropical Medicine & International Health*. 2002, 7(1): 35-44.
- Kedzierski L, Zhu Y, Handman E. Leishmania vaccines: progress and problems. *Parasitology*. 2008, 13(3)S2: 87-112.
- Reithinger R, Dujardin JC, Louzir H, Pirmez C, Alexander B et al. Cutaneous leishmaniasis. *The Lancet Infectious Diseases*. 2007, 7(9): 591-596.
- Palatnik-de-Sousa CB. Vaccines for leishmaniasis in the fore coming 25 years. *Vaccine*. 2008, 26(14): 1709-1724.
- Selvapandiyan A, Dey R, Gannavaram S, Lakhal-Naouar I, Duncan R, et al. Immunity to Visceral Leishmaniasis Using Genetically Defined Live-Attenuated Parasites. *Journal of Tropical Medicine*. 2012, 631460: 12.
- De Groot AS and Moise L. New tools, new approaches and new ideas for vaccine development. *Expert Review of Vaccines*. 2007, 6(2): 125-127.
- Gomez MA, Contreras I, Hallé M, Tremblay ML, McMaster RW et al. Leishmania GP63 alters host signaling through cleavage-activated protein tyrosine phosphatases. *Science Signaling*. 2009, 2(90): 58.
- Gómez MA and Olivier M. Proteases and phosphatases during Leishmania-macrophage interaction: paving the road for pathogenesis. *Virulence*. 2010, 1(4): 314-318.
- Webb JR, Button LL and McMaster WR. Heterogeneity of the genes encoding the major surface glycoprotein of Leishmaniadonovani. *Molecular and Biochemical Parasitology*. 1991, 48(2): 173-184.
- Olivier M and Hassani K. Protease inhibitors as prophylaxis against leishmaniasis: new hope from the major surface protease gp63. *Future Medicinal Chemistry*. 2010, 2(4): 539-542.
- Bouvier J, Etges RJ and Bordier C. Identification and purification of membrane and soluble forms of the major surface protein of leishmaniapromastigotes. *The Journal of Biological Chemistry*. 1985, 260(29): 15504-15509.
- Bouvier J, Etges R and Bordier C. Identification of the promastigote surface protease in seven species of Leishmania. *Molecular and Biochemical Parasitology*. 1987, 24(1): 73-79.
- Turco SJ, Hull SR, Orlandi PA and Shepherd SD. Structure of the major carbohydrate fragment of the Leishmania-donovani lipophosphoglycan. *Biochemistry*. 1987, 26(19): 6233-6238.

20. Medina-Acosta E, Beverley SM, Russell DG. Evolution and expression of the *Leishmania* surface proteinase (gp63) gene locus. *Infectious Agents and Disease*. 1993, 2(1): 25-34.
21. Gasim S, Elhassan AM, Kharazmi A, Khalil EA, Ismail A et al. The development of post-kala-azar dermal leishmaniasis (PKDL) is associated with acquisition of *Leishmania* reactivity by peripheral blood mononuclear cells (PBMC). *Clinical and Experimental Immunology*. 2000, 119(3): 523-529.
22. Dranoff G, Jaffee E, Lazenby A, Golumbek P, Levitsky H et al. Vaccination with irradiated tumor cells engineered to secrete murine granulocyte-macrophage colony-stimulating factor stimulates potent, specific, and long-lasting anti-tumor immunity. *Proceedings of the National Academy of Sciences of the United States of America*. 1993, 90(8): 3539-3543.
23. Jenkins MK, Khoruts A, Ingulli E, Mueller DL, McSorley SJ et al. In vivo activation of antigen-specific CD4 T cells. *Annual Review of Immunology*. 2001, 19: 23-45.
24. Germain RN. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell*. 1994, 76(2): 287-299.
25. Rosenberg ES, Billingsley JM, Caliendo AM, Boswell SL, Sax PE et al. Vigorous HIV-1-specific CD4+ T cell responses associated with control of viremia. *Science*. 1997, 278(5342): 1447-1450.
26. Rudolph MG, Stanfield RL, Wilson IA. How TCRs bind MHCs, peptides, and co-receptors. *Annual Review of Immunology*. 2006, 24: 419-466.
27. Cresswell P. Assembly, transport, and function of MHC class II molecules. *Annual Review of Immunology*. 1994, 12: 259-293.
28. Jones EY, Fugger L, Strominger JL and Siebold C. MHC class II proteins and disease: a structural perspective. *Nature Reviews: Immunology*. 2006, 6(4): 271-282.
29. Godkin AJ, Smith KJ, Willis A, Tejada-Simon MV, Zhang J et al. Naturally processed HLA class II peptides reveal highly conserved immunogenic flanking region sequence preferences that reflect antigen processing rather than peptide-MHC interactions. *Journal of Immunology*. 2001, 166(11): 6720-6727.
30. Consogno G, Manici S, Facchinetti V, Bachi A, Hammer J et al. Identification of immunodominant regions among promiscuous HLA-DR restricted CD4+ T cell epitopes on the tumor antigen MAGE-3. *Blood*. 2003, 101(3): 1038-1044.
31. Zhang GL, Khan AM, Srinivasan KN, August JT and Bruscia V. MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides. *Nucleic Acids Research*. 2005, 33(sup 2): W172-W179.
32. Hammer J. New methods to predict MHC-binding sequences within protein antigens. *Current Opinion in Immunology*. 1995, 7(2): 263-269.
33. De Groot AS, Jesdale BM, Szu E, Schafer JR. An interactive website providing MHC ligand predictions: application to HIV research. *AIDS Res Hum Retroviruses*. 1997, 13: 539-541.
34. Schafer JA, Jesdale BM, George JA, Kouttab NM, De Groot AS. Prediction of well-conserved HIV-1 ligands using a Matrix-based Algorithm, EpiMatrix. *Vaccine*. 1998, 16(19): 1880-1884.
35. Sturniolo T, Bono E, Ding J, Radrizzani L, Tuereci O et al. Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nat Biotechnol*. 1999, 17: 555-61.
36. Meister GE, Roberts CG, Berzofsky JA, De Groot AS. Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences. *Vaccine*. 1995, 6: 581-591.
37. De Groot AS, Marcon L, Bishop EA, Rivera D, Kutzler M et al. HIV vaccine development by computer assisted design: the GAIA vaccine. *Vaccine*. 2005, 23: 2136-2148.
38. De Groot AS, Levitz L, Ardito MT, Skowron G, Mayer KH et al. Further progress on defining highly conserved immunogenic epitopes for a global HIV vaccine. *Human Vaccines & Immunotherapeutics*, 2012, 8(7): 987-1000.
39. Levitz L, Koita OA, Sangare K, Ardito MT, Boyle CM et al. Conservation of HIV-1 T cell epitopes across time and clades: Validation of immunogenic HLA-A2 epitopes selected for the GAIA HIV vaccine. *Vaccine*. 2012.
40. Bond KB, Sriwanthana B, Hodge TW, De Groot AS, Mastro TD et al. An HLA directed molecular and bioinformatics approach identifies new HLA-A11 HIV-1 subtype E cytotoxic T lymphocyte epitopes in HIV-1-infected Thais. *AIDS Res Hum Retroviruses*. 2001, 17(8): 703-717.
41. McMurry JA, Gregory SH, Moise L, Rivera D, Buus S et al. Diversity of *Francisella tularensis* Schu4 antigens recognized by T lymphocytes after natural infections in humans: Identification of candidate epitopes for inclusion in a rationally designed tularemia vaccine. *Vaccine*, 2007, 25: 3179-3191.
42. Dong Y, Dimaria S, Sun X, Jesdale BM, De Groot AS et al. HLA-A2- restricted CD8+ cytotoxic T cell responses to novel *Mycobacterium tuberculosis* targets superoxide dismutase and alanine dehydrogenase. *Infect Immun*, 2004, 72(4): 2412-2415.
43. Otero M, Calarota SA, Dai A, De Groot AS, Boyer JD et al. Efficacy of novel plasmid DNA encoding vaccinia antigens in

improving current smallpox vaccination strategy. *Vaccine*. 2006, 24: 4461–4470.

44. Moise L, Buller RM, Schriewer J, Lee J, Frey SE et al. VennVax, a DNA-prime, peptide-boost multi-T-cell epitope poxvirus vaccine, induces protective immunity against vaccinia infection by T cell response alone. *Vaccine*. 2011, 29(2011): 501–511.

45. Moise L, McMurry JA, Pappo J, Lee D, Moss SF et al. Identification of genome-derived vaccine candidates conserved between human and mouse-adapted strains of *H. Pylori*. *Human Vaccines*. 2008, 4(3): 1–6.

46. De Groot AS, Ardito M, McClaine EM, Moise L, Martin WD. Immunoinformatic comparison of T-cell epitopes contained in novel swine-origin influenza A (H1N1) virus with epitopes in 2008–2009 conventional influenza vaccine. *Vaccine*. 2009, 27(42): 5740–5747.

47. De Groot AS, Ardito M, Terry F, Levitz L, Ross T et al. Low immunogenicity predicted for emerging avian-origin H7N9: implication for influenza vaccine design. *Hum Vaccin Immunother*. 2013, 9(5): 950–956.

48. Moise L, Gutierrez AH, Bailey-Kellogg C, Terry F, Leng Q et al. The two-faced T cell epitope: Examining the host-microbe interface with Janus Matrix. *Hum Vaccin Immunother*. 2013, 9(7): 1577–1586.

49. Moise L, Terry F, Gutierrez AH, Tassone R, Losikoff P et al. Smarter vaccine design will circumvent regulatory T cell-mediated evasion in chronic HIV and HCV infection. *Front Microbiol*. 2014, 5: 502.

50. Elfaki MEE, Khalil EAG, DeGroot AS, Musa AM, Gutierrez A, et al. Immunogenicity and immune modulatory effects of in silico predicted *L. donovani* candidate peptide vaccines. *Hum Vaccin Immunother*. 2012, 8 (12): 1769–1774.

51. McMurry JA, Kimball S, Lee JH, Rivera D, Martin W et al. Epitope-driven TB vaccine development: a streamlined approach using immune-informatics, ELISpot assays, and HLA transgenic mice. *Current Molecular Medicine*. 2007, 7(4): 351–368.

52. Arnold PY, La Gruta NL, Miller T, Vignali KM, Adams PS et al. The majority of immunogenic epitopes generate CD4+ T cells that are dependent on MHC class II-bound peptide-flanking residues. *Journal of Immunology*. 2002, 169(2): 739–749.

53. Rogier. Challenge of developing anti-parasite vaccines in the tropics. *Médecine Tropicale: Revue du Corps de Santé Colonial*. 2007, 67(4): 328–334.

54. Noazin S, Modabber F, Khamesipour A, Smith PG, Moulton LH et al. First generation leishmaniasis vaccines: a review of field efficacy trials. *Vaccine*. 2008, 26(52): 6759–6767.

55. Purcell AW, McCluskey J and Rossjohn J. More than one reason to rethink the use of peptides in vaccine design. *Nature Reviews: Drug Discovery*. 2007, 6(5): 404–414.

56. De Groot AS, Knopf PM, Martin B. De-immunization of therapeutic proteins by T cell epitope modification. In: Mire-Sluis, A, editor. *State of the art analytical methods for the characterization of biological products and assessment of comparability*. Dev Biol. Karger: Basel. 2005, 122: 137–160.

57. He L, De Groot AS, Gutierrez AH, Martin WD, Moise L et al. Integrated assessment of predicted MHC binding and cross-conservation with self reveals patterns of viral camouflage. *BMC Bioinformatics*. 2013, 15(Suppl 4): S1.

58. Nylén S, Sacks D. Interleukin-10 and the pathogenesis of human visceral leishmaniasis. *Trends Immunol*. 2007, 28(9): 378–384.

59. Singh OP, Gidwani K, Kumar R, Nylén S, Jones SL et al. Reassessment of immune correlates in human visceral leishmaniasis as defined by cytokine release in whole blood. *Clin Vaccine Immunol*, 2012, 19(6): 961–966.

60. Käll L, Krogh A and Sonnhammer EL. A combined transmembrane topology and signal peptide prediction method. *J Mol Biol*, 2004, 338: 1027–1036.

61. Juncker AS, Willenbrock H, Von Heijne G, Brunak S, Nielsen H et al. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Science*. 2003, 12(8): 1652–1662.

62. Fankhauser N, Mäser P. Identification of GPI anchor attachment signals by a Kohonen self-organizing map. *Bioinformatics*. 2005, 21(9): 1846–1852.

63. Meister GE, Roberts CGP, Berzofsky JA and De Groot AS. Two novel T cell epitope prediction algorithms based on MHC-binding motifs; comparison of predicted and published epitopes from *Mycobacterium tuberculosis* and HIV protein sequences. *Vaccine*. 1995, 13(6): 581–591.

64. Shannon P, Markiel A, Ozier O, Baliga SN, Wang JT et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res*. 2003,13.